Generative models for molecular discovery: Recent advances and challenges

Abstract

Development of new products often relies on the discovery of novel molecules. While conventional molecular design involves using human expertise to propose, synthesize, and test new molecules, this process can be cost and time intensive, limiting the number of molecules that can be reasonably tested. Generative modeling provides an alternative approach to molecular discovery by reformulating molecular design as an inverse design problem. Here, we review the recent advances in the state■of■the■art of generative molecular design and discusses the considerations for integrating these models into real molecular discovery campaigns. We first review the model design choices required to develop and train a generative model including common 1D, 2D, and 3D representations of molecules and typical generative modeling neural network architectures. We then describe different problem statements for molecular discovery applications and explore the benchmarks used to evaluate models based on those problem statements. Finally, we discuss the important factors that play a role in integrating generative models into experimental workflows. Our aim is that this review will equip the reader with the information and context necessary to utilize generative modeling within their domain.